Abdollahi, Sara; Nejdl, Wolfgang; Gottschalk, Simon

Retrieval-Augmented Generation of Event Collections from Web Archives and the Live Web

International Journal on Digital Libraries

Provided in Cooperation with:
Springer Nature

Suggested Citation: Abdollahi, Sara; Nejdl, Wolfgang; Gottschalk, Simon (2025) : Retrieval-Augmented Generation of Event Collections from Web Archives and the Live Web, International Journal on Digital Libraries, ISSN 1432-1300, Springer London, London, Vol. 26, Iss. 2, https://doi.org/10.1007/s00799-025-00419-7

International Journal on Digital Libraries (2025) 26:12
https://doi.org/10.1007/s00799-025-00419-7

Retrieval-Augmented Generation of Event Collections from Web Archives and the Live Web

Sara Abdollahi[1] · **Wolfgang Nejdl**[1] · **Simon Gottschalk**[1]

Received: 19 August 2024 / Revised: 16 April 2025 / Accepted: 17 April 2025 / Published online: 23 June 2025

Abstract

Creating collections of societally impactful events is a challenging task given the sheer amount of information about such events covering a large variety of aspects and perspectives in web archives and the live web. The automatic creation of such collections from web archives typically does not live up to the high standards of web archivists, who put lots of manual effort into carefully curating collections. Furthermore, the lack of engaging presentation methods sets up a burden for any users aiming to interact effectively with event collections in order to explore an event in its entirety. Therefore, we (i) conduct expert interviews to determine the requirements for building and utilising event collections from the perspectives of web archivists, (ii) introduce *EventExplorer* – a retrieval-augmented generation (RAG) approach to create event collections through efficient retrieval and diversified ranking – and make it available in an interactive web system, (iii) apply *EventExplorer* on different sources including a web archive and the live web, (iv) discuss which requirements are met by *EventExplorer* as well as the challenges that remain for future work, with a specific emphasis on the distinctive characteristics of both archived web and the live web environments. We demonstrate the effectiveness of *EventExplorer* applied on web archives through a user study of our interactive system. Then, we transfer our lessons learned to the live web by creating event collections of 166 elections in Europe. Our evaluation results show the effectiveness of *EventExplorer* in addressing the requirements identified in our expert interviews. Further, we derive a set of challenges and potential future steps for bringing together the automatic creation of web archive collections and manual curation. Finally, we discuss how to make web archives ready for their use in RAG systems.

Keywords Web archives · Event collections · Retrieval-augmented generation · Knowledge graph

1 Introduction

The world is constantly facing societally impactful events such as terrorist attacks, migrant crises, and elections, resulting in a vast amount of event-centric information on the web. Such information, subject to dynamic changes when the event unfolds, is typically told from several views and covers a large variety of aspects. Web archives, which contain the contents of websites in different snapshots over time, serve as an indispensable tool for understanding the unfolding of an event in retrospect [1].

Consequently, digital humanities researchers and social scientists rely on web archives to analyse the significant events that influence and shape our societies [2]. To facilitate the understanding and managing of the growing volume of web archives, web archivists organise them into collections [3]. An example of such endeavours is Archive-It, a service of the Internet Archive that provides manually curated collections. Fig. 1 shows an example of the Archive-It Brexit collection[1] that contains links to 71 archived websites about Brexit, manually selected by web archivists and only marginally categorised into a set of groups like "Agencies and other bodies" and "European Parliament".

The creation of such an event collection requires manual effort which is not feasible to do for all events of potential

✉ Sara Abdollahi
abdollahi@L3S.de

Wolfgang Nejdl
nejdl@L3S.de

Simon Gottschalk
gottschalk@L3S.de

[1] Leibniz Universität Hannover, L3S Research Center, Hannover, Lower Saxony, Germany

[1] https://archive-it.org/collections/11980

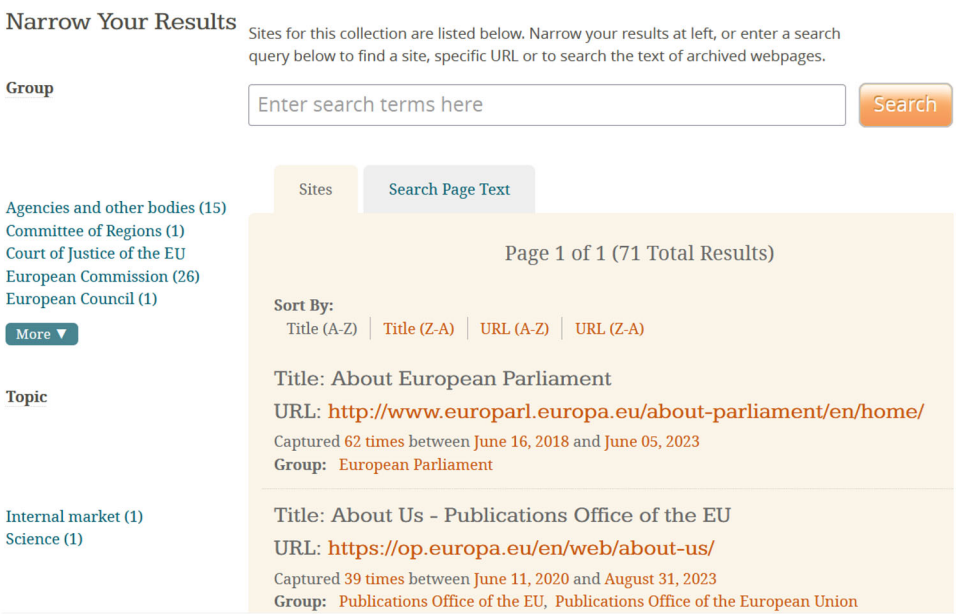


Fig. 1 The Brexit collection in Archive-It

interest to a user. Moreover, dealing with the large volume of information in web archives presents several challenges from the perspective of user interaction [4–6]. Therefore, there is a need for methodologies for the automatic creation of event collections that adapt to the specific constraints of web archives, including time-consuming access to websites [7] and the lack of dense retrieval methods [8].

In this article, our goal is to explore methods for the automatic creation of event collections from web archives and the live web, and to understand the challenges and potentials when creating event collections from different sources.

In the first step, we perform extensive interviews with experts from different European web archiving initiatives that specifically target at the creation of archived collections. Then, to perform a deeper analysis of automatically created event collections, we conduct a user study on *EventExplorer*, a new system for creating and exploring event collections from web. *EventExplorer* aims to provide the complete picture of an event through (i) identification of event aspects in a knowledge graph, (ii) efficient website retrieval through aspect-related queries, and (iii) diversified ranking techniques using the retrieved website snippets to cover a wide range of event aspects. For exploring the created event collections, we employ a large language model (LLM) via retrieval-augmented generation (RAG) [9], where we let the LLM create different descriptive components (summaries, metadata and timelines) which are interlinked with the retrieved websites.

We evaluate *EventExplorer*'s event collection on web archives and the live web: First, we perform a user study where our web archive experts explore event collections created from the Portuguese Web Archive[2]. Then, we explore *EventExplorer*'s performance on the live web using the Bing search. While the live web comes with less resources regarding the temporal evolution of an event unfolding, it provides results of high quality and relevance. This exploration helps us understand the potential of *EventExplorer* in different web environments, providing insights into how the features of the live web data impact the created event collections.

With our expert interviews and the studies of *EventExplorer*, we aim to find answers to the following research questions:

(RQ1) What are the essential requirements for both creating and utilising event collections effectively?
(RQ2) How can retrieval-augmented generation (RAG) improve user interaction and exploration of event collections?
(RQ3) How can the RAG-based event collection creation be adopted to different settings like the live web?
(RQ4) Is full automation feasible for the creation of event collections from web archives?
(RQ5) What potential research directions emerge from the findings of this study?

To address these research questions, our contributions in this article are as follows:

- We conduct interviews with experts in web archives and digital libraries to understand the requirements for building and using event collections in web archives.

[2] https://arquivo.pt/

- We propose an algorithm for creating event collections from web based on a retrieval step guided by an event knowledge graph and BERT-based ranking.
- We design and implement an interactive web system[3] that enables intuitive navigation and exploration of event collections.
- In a user study, we demonstrate how retrieval augmented generation with LLMs in our web system can help explore event collections created from the Portuguese Web Archive.
- We perform an analysis comparing our event collection approach on five sources: web archives (Portuguese Web Archive), the live web (Google News and Bing), and LLMs (ChatGPT and Microsoft Copilot).
- We perform a case study of *EventExplorer* in the live web by creating event collections of elections in Europe.

The remainder of this article is structured as follows: Next, we present related work (Section 2). Then, in Section 3, we gather insights of web archiving experts to understand the key requirements for creating and utilising event collections *(RQ1)*. In Section 4, we describe the task of event collection creation and *EventExplorer*, our approach towards it, with a specific focus on RAG (Section 4.7) *(RQ2)*. We evaluate *EventExplorer* through user studies and further analyses on archived and the live web in Section 5 and Section 6, respectively *(RQ3)*. Based on our findings, we discuss the potential of automatic creation of event collections *(RQ4)* and identify key challenges for future developments *(RQ5)* in Section 7. We conclude in Section 8.

2 Related Work

In this section, we review the existing works in four key areas relevant to our study.

2.1 Search and Exploration in Web Archives

Various studies have delved into understanding user needs, examining archivists' practices, and identifying challenges in engaging users in the domain of web archiving. Costa et al. [10] explore the information needs of web archive users and their search behaviours by gathering feedback from them. In another work [11], the authors observe web archivists at the Internet Archive using an ethnographic approach to understand their practices. Hockx-Yu [12] analyse the factors contributing to the limited scholarly use of web archives and formulated a set of requirements for improving them. Jones et al. [13] investigate the temporal nature and the processes involved in the creation of collections by Archive-It web archivists. Jayanetti et al. [3] reviewe eight web archive platforms' collection structures and navigational hierarchies.

Several studies have focused on storytelling to tackle the challenges of exploring large volumes of web archive collections and to enhance users' experience. AlNoamany et al. [5] identify candidate web pages within a collection and employ tools like Storify to summarise and present them to users. Jones et al. [14] introduce MementoEmbed a tool designed to summarise collections and generate stories, facilitating exporting these narratives for archivists. Furthermore, Hypercane [15, 16] is a tool that utilises storytelling techniques to select exemplar pages from web archive collections and display them in a social media interface. Jones et al. [17] study the role of surrogates in understanding a web archive collection by users. [18] improve collection understanding by summarising these collections and visualising the summaries.

Regarding developing interfaces for web archive exploration, Archives Unleashed Cloud [19] offers a web-based interface for working with web archives at scale. Kelly et al. [20] propose a method for users to navigate personalised representations of mementos. Furthermore, Public Archive [21] offers an intuitive interface for users to search and analyse a vast collection of preserved news articles from the Portuguese reference newspaper *Jornal Público*.

2.2 Event Collections and Narratives

In the ongoing effort to create web archive collections specifically for events, Nanni et al. [22] use relevant concepts and entities collected from a knowledge base. They extract a set of potentially relevant entities, collect some related passages, and finally use embedding techniques in a learning to rank method to provide a final rank list of documents. Gossen et al. [23] present an approach based on focused crawling of web archives to create event-centric collections without requiring any full-text indices. Demidova et al. [1] explore techniques for constructing event-centric collections from large-scale web archives, focusing on crawl-based methods that identify relevant documents and incorporate link networks as contextual elements. Plötzky et al. [24] explore narratives that enhance the semantics of complex real-world events and describe their internal structure from multiple perspectives. The authors propose a narrative model that recursively represents narratives related to real-world events.

2.3 Query Expansion and Dense Ranking

In recent years, large language models (LLMs) have significantly improved document ranking. Nogueira et al. [25] first applied BERT [26] to document ranking, followed by several methods [27–30] utilising LLMs for document ranking. Query expansion is a technique aimed at enhancing retrieval

[3] https://eventexplorer.l3s.uni-hannover.de/

performance by adding terms to the user's query. Various approaches leverage knowledge graphs to find related entities to the query [31–35] for query expansion. Two approaches focus specifically on event-related queries: Rosin et al. [36] utilise Wikipedia2Vec embeddings and Abdollahi et al. [37] use EventKG [38] and BERT [26].

2.4 Retrieval-Augmented Generation (RAG)

Retrieval-Augmented Generation (RAG) [9] facilitates text generation based on information retrieved from external sources. This approach effectively addresses the challenges of LLM "hallucinations" and access to the most recent data. RAG has been applied across various domains such as language modelling [39, 40], text summarization [41], paraphrase generation [42] and citation identification [43].

Recent studies have expanded RAG's applicability to more specialised areas. Ji et al. [44] introduce the RAG-RLRC-LaySum framework, which uses RAG to generate comprehensible text from biomedical scientific content for non-specialists, making complex information more understandable. Similarly, He et al. [45] implement a RAG approach to generate headlines and important numerical data from news articles through an end-to-end training pipeline. Furthermore, Zhang et al. [46] propose REACT, which generates commit messages related to automated software engineering tasks.

3 Requirements Analysis through Expert Interviews

We conducted extensive interviews with web archive experts to identify the requirements of event collections created from the rich contents available in web archives.

Our interviews are guided by the following three goals:

- *(G1)* What is the process of creating and maintaining event collections?
- *(G2)* What are user needs and expectations of web archive collections?
- *(G3)* What are the characteristics of event collections?

3.1 Setup & Participants

We conducted online video interviews with web archive curators and web archivists, all employed by different national web archives in Europe. They were selected for their extensive expertise in the field of web archiving: Each expert rated their knowledge of web archives, web archive collections and their creation and maintenance as "advanced" or "expert". Further, they dedicated significant time to in-depth interviews, providing valuable insights: The five interviews to identify requirements took about 90 minutes each. Further, after we built our interactive web systems based on the identified requirements, the experts have been available for our user study as described in Section 5.3. This way, we ensured high-quality insights by engaging experts with extensive experience in creating web archive collections. Their deep understanding of technical requirements and user needs guided our research and development process from start to finish.

The interview questions covered key areas related to web archive event collections, including the selection criteria for topics, creation processes, user information needs, collection characteristics, and specific features associated with event collections. To guide our interviews, we prepared 18 questions listed in the first column of Tables 7, 8 and 9 in the Appendix.

3.2 Findings

Tables 7, 8 and 9 in the Appendix provide a summary of our interviews. We identify the following common patterns in the responses provided by various experts and institutions regarding the creation of web archive collections:

Curators' expertise *(G1)*: Effective event collection curation requires understanding the event and its key aspects. However, finding experts with the necessary knowledge for event-specific curation is typically challenging.

Resource and time constraints *(G1)*: Traditional collection methods are resource-intensive and time-consuming.

Managing information overload *(G1)*: Event collections distinguish themselves from thematic collections by their temporal aspect. Events happen at specific moments in time, resulting in an explosion of online information. Handling a huge amount of information given the limited time and resources is crucial in collection creation.

Users *(G2)*: While assumptions are made that users are from the general public and academia, there is still a lack of knowledge and studies about web archive users and their information needs.

Quality *(G3)*: A successful event collection covers key aspects of the event and connects with users, providing valuable insights and information.

Size and presentation *(G3)*: Web archive collections can range from as few as 50 to several thousand websites. Size and display are two important factors in user engagement. Well-organised collections with rich metadata and sub-categories facilitate user navigation and improve user engagement.

Reflecting time and society *(G3)*: Web archive collections reflect the context of their time and are considered the collective memory of a country (or community). They are valuable resources, documenting the narrative of events and the evolution of thoughts and views of people in a time.

3.3 Requirements

From our interview findings, we derive the following requirements for the creation of event collections from web archives *(RQ1)*:

Coverage: A successful event collection should comprehensively cover key aspects of the event, establishing meaningful connections with users by offering valuable insights and information.

Structure: Collections must be well-organised, include rich metadata, sub-categories, and carefully selected websites to improve overall accessibility. Sub-collections and sub-categories can help meet user information needs and their navigation and enhance overall accessibility as the size of collections increases.

Plurality: To ensure representation, event-centric collections approaches should capture the voices on the margins and people with diverse backgrounds and views (e.g., gender, politics, ethics, religion).

While *plurality* is a highly relevant requirement for the representativeness of an event collection, it requires to discover and understand viewpoints in the documents which is a highly complex and potentially subjective task [47, 48]. We address this requirement to some extent by incorporating diversity-based ranking, but we primarily focus on the first two requirements in the remainder of this article. Also, we focus on text documents and not other modalities (e.g., videos) or social media.

4 Event Collection Creation

In this section, we first introduce the notation of our model to build event collections from web[4]. Then, we introduce *EventExplorer*, our approach to automatic event collection creation based on our insights gained from the interviews (Section 3).

4.1 Problem Statement

In this article, we focus at the task of event collection creation.

Definition 1 (Event Collection Creation) Let D be the text documents[5] on the web and EKG an event knowledge graph that contains events, entities and their relations [38, 49]. Given a query event of interest q, the goal of event collection creation is to return a ranked set of documents $D_q = \{d_1, d_2, ..., d_k\}$ which represent the coverage of the event on the web.

[4] In this context, web refers to both web archives and the live web.

[5] Throughout this article, we use the terms "document" and "website" interchangeably.

Following the requirements identified in our interviews, we require the generated event collections to be structured through sub-collections [4]:

Definition 2 (Event Collection Structuring) Event collection structuring represents the event collection as a set of sub-collections $S_q = \{D_{q,1}, D_{q,2}, ..., D_{q,k}\}$ where each sub-collection $D_{q,i} \in S_q$ contains documents relevant to a specific aspect $aspect_i$ of the event.

The goal of event collection structuring is to represent the event collection in a way to make it accessible to a user for exploring the event.

For the event collection and each of its sub-collections, a set of descriptive components enables navigation through the collections:

Definition 3 (Descriptive Components) For each event (sub-)collection and its documents D, a set of *descriptive components* needs to be generated: (i) a *summary* briefly describing the event based on D, (ii) *metadata* (time, participants, locations) extracted from D and (iii) a *timeline* denoting important happenings reported in D. For summary and timeline, references to the specific documents in D from which information is extracted need to be provided.

These components allow a user to interact with the collection and its documents efficiently.

4.2 Overview of *EventExplorer*

Fig. 2 gives an overview of our approach *EventExplorer*[6] at the example of using a web archive for event collection generation. *EventExplorer* identifies event aspects to capture user information needs (structure requirement). For each aspect, *EventExplorer* retrieves documents from a web archive's API. The documents are ranked with the *monoBERT* ranking model [25] and diversified following the coverage requirement identified in the interviews. A general ranking of documents for the event is achieved by aggregating all aspects. For user interaction, *EventExplorer* follows the naive Retrieval-Augmented Generation (RAG) approach [9] where we use the retrieved document collections to augment the prompts to a Large Language Model (LLM) for generating three types of components describing the events.

It is important to note that *EventExplorer* is capable of adapting to the specific constraints of web archives. First, due to the dimensionality of web archives, there is no possibility of creating our own indexes over the documents D. Instead, *EventExplorer*'s document retrieval step can only

[6] The *EventExplorer* retrieval and ranking code are available online: https://github.com/saraabdollahi/EventExplorer

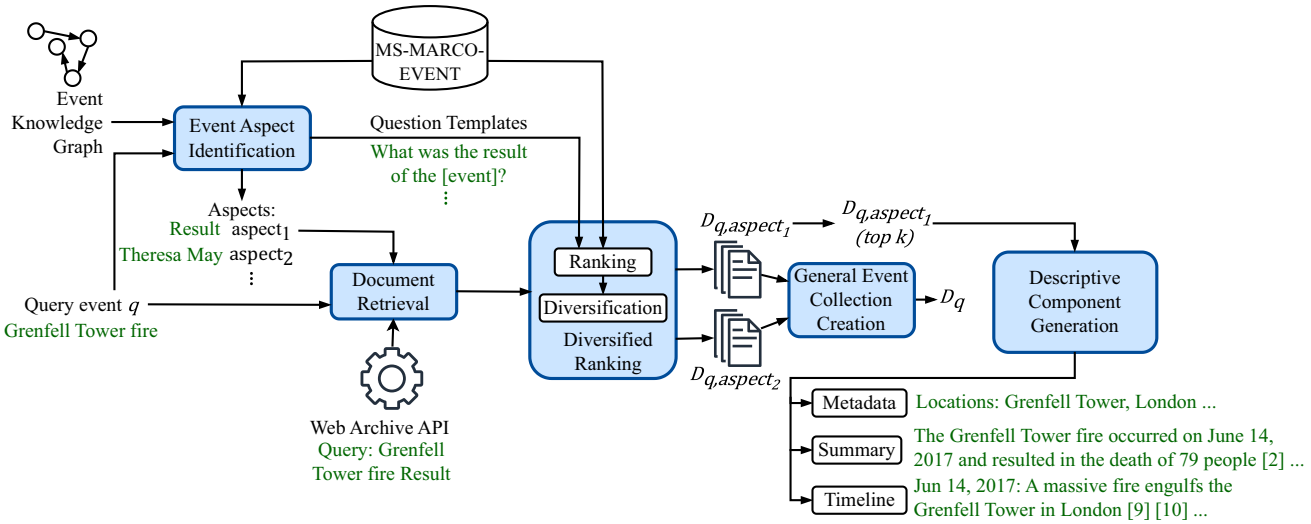


Fig. 2 Overview of *EventExplorer* applied on a web archive. An example event ("Grenfell Tower fire") and two of its aspects ("Result" and "Theresa May") are shown in green. The generation of three descriptive components for the documents in the first aspect is shown

access documents by boolean term-based search using existing web APIs[7]. Consequently, *EventExplorer* allows event collection creation in an ad-hoc manner and can easily be extended to different web archives with search functionalities. Second, retrieving the full content of archived websites is typically time-consuming [7]. Therefore, *EventExplorer* only requires the availability of website snippets, typically retrieved together with the query results.

4.3 Event Aspects Identification

Our approach to creating collections goes beyond the relevance of individual documents to queries. Considering documents in isolation and independent of other documents, we may encounter the potential issue of retrieving highly relevant yet redundant information. To avoid this issue, *EventExplorer* considers *aspects* as representations of users' information needs (similar to the concept of information nuggets [51]), ensuring a more comprehensive and diverse understanding of the event. Each aspect models specific details that users seek regarding an event.

4.3.1 Event Aspects Analysis

In a first preprocessing step, we aim to uncover the diverse information needs of users regarding events. Precisely, we manually explore a collection of event-based user queries and find out what specific aspects of events (e.g, the event date and location) users are interested in. Further, we analyse if these aspects differ across event types (e.g., sports events vs. crises). The derived event aspects statistics will later help us in retrieving relevant aspects regarding the query event of interest.

To conduct our event aspects analysis, we leverage the $MS-MARCO_{EVENT}$ dataset[8], which consists of user queries related to events (e.g., "What day was the Galveston hurricane?"). As event types, we use classes in the DBpedia ontology (e.g., `SportsEvent`[9]). We have manually annotated the queries in $MS-MARCO_{EVENT}$ to identify a set of recurring aspects and their distributions across event types.[10]

Our analysis reveals that there are five frequently recurring aspects across event types (similar to the event elements in [52]): *When*, *Where*, *Who*[11], *Result* and *Cause*. To illustrate these aspects with an example, Table 1 presents one selected example query per aspect from $MS-MARCO_{EVENT}$ for the event *World War II*. The distribution of event aspects for two specific event types is shown in Table 2.

Only 12% of the queries in $MS-MARCO_{EVENT}$ could not be assigned to any of these five event aspects. Examples include the queries "Why is it called George Pickett's migration at the Battle of Gettysburg?" and "What was the major difference between the American and French revolutions?".

[7] Examples of such APIs that allow full-text search in web archives include the Wayback Machine (https://archive.org/advancedsearch.php) and the Portuguese Web Archive (https://arquivo.pt/) [50].

[8] https://github.com/saraabdollahi/MusQuE/tree/main/Data

[9] https://dbpedia.org/ontology/SportsEvent

[10] We manually performed event aspect identification as a one-time preprocessing step to ensure accuracy. However, we acknowledge that text classification could be adapted for automation.

[11] Given a query event, we will identify more fine-grained aspects of *Where* and *Who*: for example, "Theresa May" is an aspect of the Brexit covering the aspect *Who*.

Table 1 One example query per aspect in the $MS-MARCO_{EVENT}$ queries about the event World War II

Aspect	Example Query
Result	How many people died during WWII?
Cause	Why was the failure of treaty of Versailles a cause of WWII?
When	What date did World War two end?
Where	Where did WWII begin?
Who	Who was leader of Italy during WWII?

Table 2 Distribution of aspects for two event types in $MS-MARCO_{EVENT}$

Aspect	`SportsEvent`	`Crisis`
When	46%	24%
Where	29%	7%
Who	15%	13%
Result	3%	31%
Cause	2%	24%
Others	5%	1%

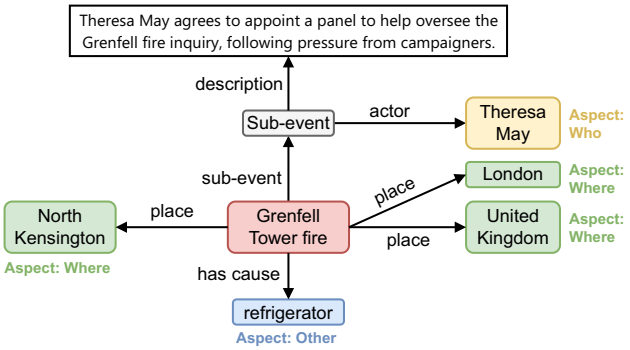


Fig. 3 Knowledge graph representation of "Grenfell Tower fire", illustrating the related places, participants, and description. Each entity is colour-coded to represent its respective aspect: green for "where", yellow for "who" and blue for "other"

Since these cases appear infrequently, we decided to focus on the mentioned five aspects for event aspects retrieval.

As a result of our event aspects analysis, we manually group the DBpedia event types into *planned events* (e.g., `SportsEvent` and `Election`) and *unexpected events* (e.g., `Crisis`). For the event aspects retrieval described in the next section, we consider *Cause*, *Result*, *When*, *Where* and *Who* as relevant aspects of unexpected events while the aspect *Cause* is not applicable for planned events (for example, as shown in Table 2, only 2% of queries about sports events ask about their cause).

4.3.2 Event Aspects Retrieval

Given the query event q, we identify relevant aspects based on the knowledge gained in our event aspects analysis. First, we select the aspects *Result* and *When*, as well as *Cause* for unexpected events. Second, we retrieve more fine-grained aspects of *Where* and *Who*, namely, the event locations and participants. To this end, we utilise the event knowledge graph EKG and select entities directly connected to q via the properties `hasPlace` and `hasActor`[12] following our previous work [37]. Entities from the EKG that are neither of type person nor location are categorised under the aspect *Other*.

As an example, Fig. 3 illustrates selected entities related to the event "Grenfell Tower fire" in an EKG connected through various relations such as place, actor, sub-event, and cause[13]. Precisely, we retrieve the aspects *Result*, *Cause*, *When*,the participants (*Who*) "Theresa May", "Metropolitan Police", "London Fire Brigade", and the locations (*Where*) "United Kingdom", "London" and "North Kensington".

Each aspect comes with its name that is later used for query expansion and a question template (e.g., "What was the result of the [event]?" and "[event] [participant]") later used for ranking.

4.4 Document Retrieval

For each identified aspect of the event, we create an expanded query where we add the aspect name to the event name (e.g., "Grenfell Tower fire London Fire Brigade"). Date filters are added to the queries following our insights from the interviews (see Table 9: "How is the time span for an event collection selected?" – answers of experts 3 and 4): for planned events, we search for documents crawled 2 months before and after the event date. For unexpected events, we consider 4 months after the event. Under these constraints, we retrieve the top 1,000 documents for each aspect from the target web archive's search interface[14], which usually employs a fast, sparse retrieval method.

[12] These are properties in the namespace of the Simple Event Model [53], which is, for example, used in the EventKG [38].

[13] To increase readability, we show property labels instead of their URIs.

[14] In the case of using the live web APIs, such as the Bing API, we follow the same retrieval process.

It is important to highlight that, to maintain efficiency, our focus during this step is specifically on retrieving document snippets instead of the full website contents.

4.5 Diversified Ranking

For each aspect, we employ the *monoBERT* ranking model [25] to rank the documents retrieved in the previous stage, where we formulate the queries to resemble natural language questions using the question templates introduced in Section 4.3.2. *monoBERT* is an efficient and robust document ranking model that has been widely adopted for document ranking and has been shown to outperforms traditional and lightweight approaches while remaining computationally efficient [54]. It has achieved significant improvements over traditional document ranking models by leveraging BERT's understanding of document and query semantics.

Building upon BERT and other language models, models such as DuoBERT [30], PARADE [29], and ColBERT [55] introduce mechanisms like passage representation aggregation, multi-stage ranking, and late interaction to improve ranking effectiveness. While these models aim to optimize efficiency, they often require additional training and involve increased computational complexity during both training and inference phases. We select *monoBERT* due to its simplicity, reproducibility, and effectiveness which make it a preferred alternative to other document ranking models built upon similar architectures.

To adapt *monoBERT* to event-specific characteristics, we have fine-tuned BERT on $MS{-}MARCO_{EVENT}$ with reformulated questions using the annotated aspects and question templates.[15]

4.5.1 Diversification

To ensure broad coverage of topics and inclusion of different views in the top-ranked documents per aspect, we perform diversified ranking through a greedy re-ranking algorithm as in [56]. First, we establish an initial ranking according to individual relevance to the query computed by *monoBERT* ($Score_{monoBERT}$). Subsequently, starting from the top-ranked document and moving down, we calculate each document's temporal and textual dissimilarity to all documents ranked above. This way, we not only consider the individual relevance of a document but also its temporal and textual distinctiveness to other highly ranked documents. The final diversified ranking score of a document d is then calculated through a linear combination of the *monoBERT* score and the average textual dissimilarity and temporal deviations.

$$\begin{aligned} Score_{diversified}(d, D_{q,i}) = \\ \alpha \cdot Score_{monoBERT}(d, Aspect_i) \\ + \beta \cdot Diversity_{text}(d, D_{q,i}) \\ + \gamma \cdot Diversity_{time}(d, D_{q,i}), \end{aligned} \tag{1}$$

where $D_{q,i}$ is the sub-collection of event q and its aspect $aspect_i$.

$Diversity_{text}$ is the degree of textual dissimilarity between a document and those ranked above measured by averaging over the cosine distances of their *monoBERT* embeddings. For a document d at rank k, i.e., d_k, it is computed as follows:

$$\text{Diversity}_{\text{text}}(d_k, D_{q,i}) = \frac{1}{k-1} \cdot \sum_{r=1}^{k-1} (1 - \text{CosineSimilarity}(d_k, d_r)) \tag{2}$$

where the $CosineSimilarity$ of two documents d_r and d_j measures their textual similarity based on their embeddings:

$$\text{CosineSimilarity}(d_r, d_j) = \frac{\text{Embed}(d_r) \cdot \text{Embed}(d_j)}{\|\text{Embed}(d_r)\| \cdot \|\text{Embed}(d_j)\|} \tag{3}$$

$Diversity_{time}$ is the mean temporal deviation between a document and the documents ranked above – measured by averaging over their differences in crawling or publication dates as follows:

$$\text{Diversity}_{\text{time}}(d_k, D_{q,i}) = \frac{1}{k-1} \cdot \sum_{r=1}^{k-1} (\text{DateDifference}(d_k, d_r)) \tag{4}$$

where $\text{DateDifference}(d_k, d_r)$ is the absolute difference in days between the two documents' crawling or publication dates.

4.6 General Event Collection Creation

After diversified ranking for each aspect, we also create a general ranking of the event documents that represents all considered aspects together, ensuring a comprehensive collection with wide coverage. To achieve this, we take all documents in each sub-collection and rank them by their respective maximum score in the sub-collections.

4.7 Descriptive Component Generation

[15] While our ranking model is trained on $MS-MARCO_{EVENT}$, it can be used to rank any documents and is not constrained on this dataset.

Fig. 4 Screenshot of the *EventExplorer* interface showing the event "Grenfell Tower fire" given the aspect "Where" and related entity "United Kingdom"

As evidenced by our findings in the interviews (Section 3.2), a major challenge when dealing with web archive collections is to manage the information overload, i.e., to enable users to traverse the collections efficiently, specifically in the context of close and distance-reading views [6]. Therefore, as introduced in Definition 3, we aim at creating the following types of descriptive components allowing intuitive exploration of the event and its documents: summary, metadata, and timeline.

For each aspect and the general ranking, we prompt an LLM three times to generate the components. To write these prompts, we use prefix prompts created through manual template engineering as described by Liu et al. [57]. To provide information about the documents to the LLM, we use RAG by augmenting the prompts with the retrieved and ranked documents.

Listing 1 shows an example prompt for generating a timeline, metadata and summary of the event "Grenfell Tower

Listing 1 Example prompt for timeline generation of the event "Grenfell Tower fire"

```
Follow my instructions as precisely as possible. Only provide the
  requested output, nothing more. I am giving you the name of and
  event and a JSON list of news articles about the event. Your task is
  to provide a timeline JSON of all the articles that should focus on
  the event. You can use all articles. For each item in the timeline,
  provide references to the articles using their identifiers. Only use
  information that is contained in the articles.

This is an example:

=== Event ===

Event: 2017 London Bridge attack

=== News Articles ===

{"11":{"Title":"Twilight Language: London Bridge Attack", "Snippet
  ":"Saturday, June 03, 2017 London Bridge Attack A white van
  mowed down pedestrians as it sped down London ...", "Crawl Date
  ":"2/August/2017"},...}

=== Timeline ===

[{"Date": "2017/June/3", "Text": "A white van mows down
  pedestrians on London Bridge and in Borough Market in a terrorist
  attack.", "Articles": ["1", "5"]}, ...]

=== End of Timeline ===

This is your task:

=== Event ===

Event: Grenfell Tower fire

=== News Articles ===

{"1": {"Title": "The View from Cullingworth: regulation", "Snippet":
  " ... Wednesday morning I started watching Grenfell Tower burning
  ... ", "Crawl Date": "5/August/2017"},...}

=== Timeline ===
```

fire" based on its general document ranking[16]. These prompts consist of an instruction followed by an example and the target event. The example consists of an event name ("2017 London Bridge Attack"), a JSON object of 20 news articles (with title, crawl date and snippet) and another JSON object with an example timeline, where each timeline entry comes with a list of identifiers (here, 1 and 5) that refer to documents related to the timeline entry. Then, the name of the target event and a JSON object of retrieved news articles is given.

With similar prompts, we generate summaries and metadata (information regarding the date range, locations and relevant subjects). We follow this approach for the general ranking and for each aspect, where we provide 10 documents and ask the LLM to specifically focus on the aspect. Table 3 shows the prompts and example outputs of the RAG-based component generation for the event "Grenfell Tower fire", including the common instruction.

4.8 *EventExplorer* Configuration

We fine-tune *monoBERT* on $MS-MARCO_{EVENT}$ utilising the `bert-base-uncased` model with 12 layers on 5 epochs. To balance between ranking score, text and time diversity, we set parameters α, β and γ in Equation 1 to 0.7, 0.15 and 0.15, respectively (more details in Section 5.2). To create descriptive components, we use the `gpt-3.5-turbo-instruct` model from OpenAI[17]. As our event knowledge graph EKG, we leverage EventKG [38] for event aspects retrieval described in Section 4.3.2.

5 Studies on Web Archives and News Collections

In this section, we study *EventExplorer*'s performance on creating event collections based on archived information. First, we evaluate *EventExplorer*'s ranking capability compared to a set of baselines. Second, we study the influences of its ranking parameters. Finally, we conduct a user study on *EventExplorer*'s results from web archives.

5.1 Ranking Evaluation

We compare *EventExplorer*'s ranking capabilities with three baselines using the gold standard introduced by Nanni et al. [22] that provides manual relevance labels of event-related articles in the NYT corpus – a corpus with over 1.8 million New York Times news articles from 1987 to 2007, covering several societal, political, and cultural events. Since this gold standard does not concern retrieval or reward diversity, we solely evaluate the ranking performance of *monoBERT* (excluding the text and time diversities) trained on $MS-MARCO_{EVENT}$ as applied by *EventExplorer*.

We compare the performance of our ranking approach with three baselines:

- Event-name [58] retrieves documents by TF-IDF cosine similarity and based on the keywords of the event name.
- IDCM [59] leverages an intra-document cascade ranking approach, employing knowledge distillation during training and utilising DistilBERT [60] architecture. It is trained on the *MS-MARCO* dataset [61].
- Nanni et al. [22] build event collections from large datasets by retrieving core documents related to an event and additional documents describing related aspects by

[16] This prompt was shortened and reformatted for brevity. For examples of complete prompts used in *EventExplorer*, see https://eventexplorer.l3s.uni-hannover.de/resources.

[17] https://platform.openai.com/

Table 3 Prompts and results of RAG-based component generation for the example event **Grenfell Tower Fire**. Each task description begins with the following common instruction to emphasise the importance of following the provided guidelines precisely: "Follow my instructions as precisely as possible. Only provide the requested output, nothing more. I am giving you the name of an event and a JSON list of news articles about the event"

Task	Prompt (Task Description)	Example Output
Summary Generation	Your task is to provide a summary of all the articles that should focus on the event. Provide references to the articles using their identifiers. Only use information that is contained in the articles.	The Grenfell Tower fire occurred on June 14 and resulted in a large number of casualties and injuries. It was treated as a major incident, with 58 assumed dead. The fire was believed to be caused by a malfunctioning appliance and the building's lack of proper fire safety measures
Metadata Generation	Your task is to provide a metadata JSON which consists of the date range of the event reported in the articles, the event locations mentioned in the articles and the event subjects mentioned in the articles.	• Time range: Jun 14, 2017 - Aug 06, 2017 • Subjects: Fire, Housing, Poverty, Safety • Locations: London, Grenfell Tower
Timeline Generation	Your task is to provide a timeline JSON of all the articles that should focus on the event. For each item in the timeline, provide references to the articles using their identifiers. Only use information that is contained in the articles.	• Jun 14, 2017: A fire engulfs the 24-floor Grenfell Tower in West London, resulting in multiple fatalities and injuries • Jun 15, 2017: Residents are asked to shine torches and phones out of their windows to assist the fire brigade in their response to the Grenfell Tower fire. • Jun 17, 2017: British pop star Adele shows her support for victims of the Grenfell Tower fire.

Table 4 Ranking performance comparison by mean average precision (MAP) for all events and three specific event types

	All	Election	War	Crisis
Event-name [58]	0.35	0.30	0.47	0.29
IDCM [59]	0.56	0.56	0.68	0.49
Nanni et al. [22]	0.64	0.54	0.75	**0.66**
EventExplorer	**0.70**	**0.72**	**0.75**	0.64

identifying. This is achieved through the identification of relevant entities from knowledge graphs.[18]

Table 4 presents the ranking performance of *EventExplorer* and the baseline methods, measured by mean average precision (MAP). For overall ranking performance across all events, *EventExplorer* achieves a MAP of 0.70, outperforming the other methods. Specifically, *EventExplorer* is superior to the baselines in "Elections" and "War" event types, achieving the highest MAP scores. For "Crisis" events, while Nanni et al. achieves the highest score, *EventExplorer* follows closely in second place. These results highlight the effectiveness and reliability of *EventExplorer* in ranking event-related documents across different categories.

[18] While re-implementing this baseline for our experiments, we have been in conversation with its authors to double-check the validity of our implementation.

In contrast to the baselines, our approach is specifically designed to leverage snippets instead of full websites to achieve greater efficiency. By employing LLMs, we can comprehend the full context of a website from just a snippet, making our method both effective and resource-efficient. In contrast, Nanni et al. rely on TF-IDF term frequencies. This traditional method can be less effective with snippets, as they may lack the context provided by full texts, potentially missing important terms dispersed throughout a longer document. Therefore, our LLM-based method offers a significant advantage in maintaining contextual understanding while enhancing efficiency.

5.2 Parameter Study

We investigate the influence of the parameters introduced in Equation 1 that control ranking diversity. Fig. 5 demonstrates the relationship between the parameters α, β, and γ and their respective impacts on various diversities and ranking scores within the top-10 highest ranked documents (the number of documents selected for descriptive components). Here, *text diversity* is the number of unique words normalised by collection size, *time diversity* is the normalised mean pairwise difference of document dates, and *mean ranking score* is the average of the documents' *monoBERT* scores.

The text and time diversity demonstrate a similar pattern of gradual increase as β and γ increase from 0.0 to 0.9. A consistent upward trend in mean ranking scores is observed

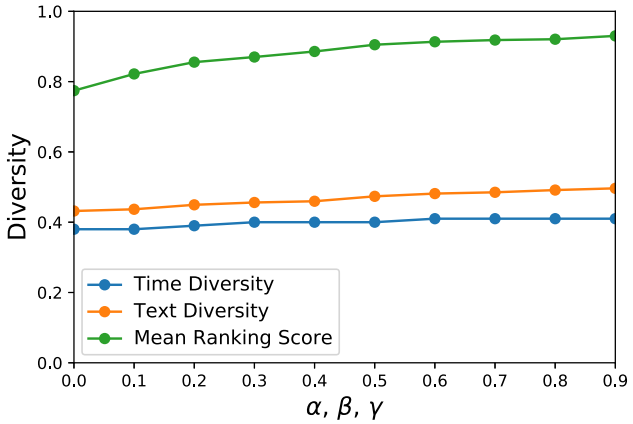


Fig. 5 Parameter study: Effects of changing the parameters α (weight of ranking score), β (weight of text diversity) and γ (weight of time diversity) of the diversified ranking (Equation 1)

as α increases, although this trend becomes less considerable beyond an α value of 0.7. This suggests that we can reach a trade-off between relevance and diversity (text and time diversities) with our configuration $\alpha = 0.7$, $\beta = 0.15$ and $\gamma = 0.15$.

5.3 User Study

To obtain feedback on automatically created event collections from web archives, we conduct a user study on *EventExplorer* with the same experts as in our interviews.

5.3.1 User Interface

To combine our approaches for document retrieval, ranking and descriptive component generation, we implemented a user interface shown in Fig. 4 that is also available online.[19]

For the purposes of the user study, *EventExplorer* retrieves English documents from the Portuguese Web Archive (PWA)[20], an openly available web archive that contains archived websites from 2007 on and serves as a valuable repository of both global and Portuguese-specific online content. Specifically, it provides a keyword-based search engine [50] used for our document retrieval step when using web archives.

5.3.2 Setup

Our user study aims to gather insights and feedback on the created event collections and user interface. It involves the five experts described in Section 3.1 who were already interviewed. For our user study, they are given a link to the *EventExplorer* website, asked to explore one or more events and to fill out an online questionnaire.

We made the following events available on the *EventExplorer* website for our user study:

- 2017 Welsh local elections
- 2017 Scottish local elections
- 2018 United Kingdom local elections
- 2020 British Grand Prix
- 2017 London Bridge Attack
- Grenfell Tower fire
- 70th British Academy Film Awards
- Poisoning of Sergei and Yulia Skripal
- Eurovision Song Contest 2018
- 2017 Catalan independence referendum

We have selected these events to ensure representation of different event types, including sports events, political events, and cultural events.

The questionnaire consists of three main sections: (i) the top websites of an event collection's general ranking, (ii) the event aspects, and (iii) the descriptive components. In each section, the participants are invited to provide additional comments. The questions are presented as statements with clickable scales, requiring the participants to indicate their agreement level, ranging from *strongly disagree* to *strongly agree*. We used Fleiss' Kappa [62] as a metric to measure the level of agreement among participants in our study, resulting in a value of $\kappa = 0.35$.

5.3.3 Results

The results of our questionnaires are presented in Fig. 6. The statements regarding the descriptive components received the highest scores, suggesting high satisfaction with summaries, metadata, and timelines. Similarly, the statement group "Event Aspects" received positive feedback. In contrast, the statements about top websites received lower scores, indicating potential limitations with document retrieval and ranking of web archives documents.

We also asked the experts to rate the usefulness of aspect types, using ratings ranging from *Not Useful* to *Extremely Useful*. Based on the usefulness scores depicted in Fig. 7, *Cause* has the highest level of usefulness, followed by *When*. The aspects *Result*, *Where*, and *Who* all have identical scores, while the aspect *Other* shows the lowest score.

Through the comments provided by the experts, we gained closer insights that are detailed in the following.

Relevance and coverage of top Websites. The experts discussed the importance of prioritising the websites originating from the country where the event occurred. They also observed that some websites lacked relevance or only a small

[19] https://eventexplorer.l3s.uni-hannover.de/user_study_pwa_en/

[20] https://arquivo.pt/?l=en

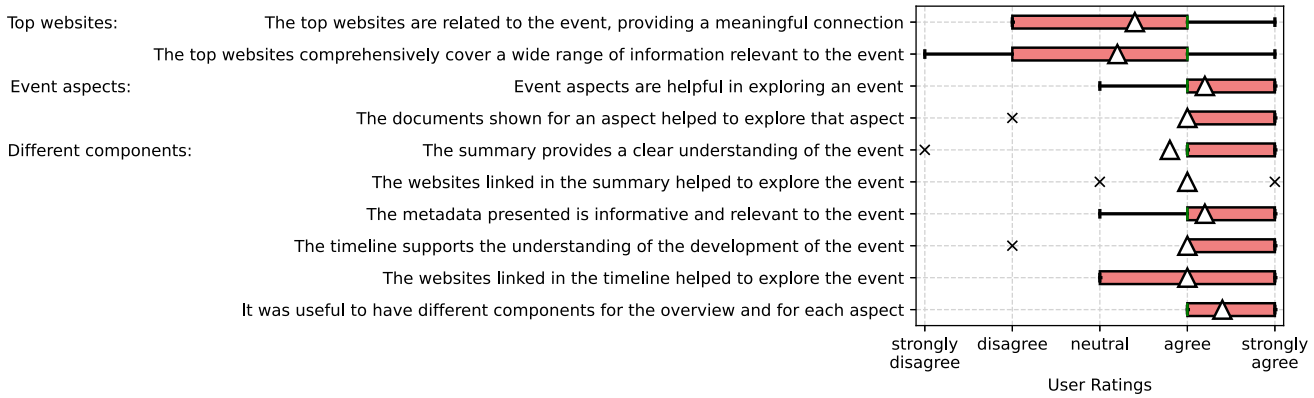


Fig. 6 Box plots of participant responses on the questionnaire statements. Δ marks the averages and × marks outliers

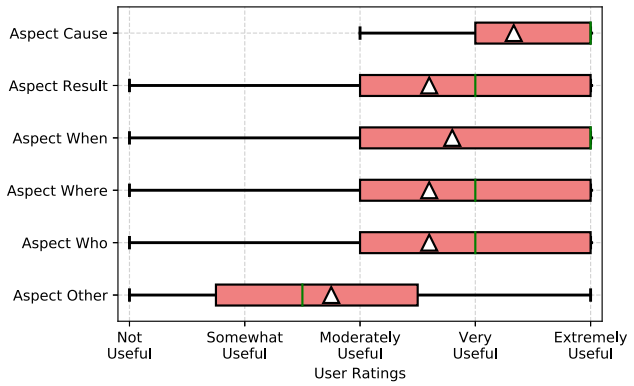


Fig. 7 Box plots of participant responses on the usefulness of event aspects. Δ marks the averages and × marks outliers

section being relevant to the event. Additionally, the participants emphasised the need for a diverse range of website types, including news media and official publications such as government statements. However, one participant highlighted that despite the limitations of relying only on English results of PWA, which represent only a small portion of the available content, our results were diverse, and the content was aggregated and structured in an interesting way.

Event Aspects. One of the participants suggested that the aspects *When* and *Where* could be adequately described in the collection's metadata. Regarding the *Result* aspect, another expert expressed difficulty in interpreting its purpose, mentioning the complexity of defining event outcomes within this framework. Overall, the experts acknowledged the event aspects, especially *Who*, *When* and *Where* as helpful, indicating their importance in the process of creating event collection and afterwards, its exploration.

Different Components. As evidenced by the final statement in Fig. 6, the experts agreed that having different summaries, metadata, and timelines for both the general ranking and each aspect was useful *(RQ2)*. According to their comments, they found the summary and its links to documents in the collection helpful, though some noted a few inaccuracies in the provided links. They noted the clarity and simplicity of the metadata but recommended expanding it to include details about the collection's provenance, such as its curators and curation process. The timeline received positive feedback for providing a clear narrative, improving understanding and navigating, especially for visual learners. Suggestions for improvement included refining the linking mechanism and providing clearer explanations of the covered time periods.

Interface Design. One participant acknowledged the value of offering different entry points into the event collections to support users with diverse learning preferences. The experts also praised the interface for effectively guiding users without biases or obstacles and for being user-friendly.

6 Studies on the Live Web

In this section, we study how *EventExplorer* performs when applied on the live web. First, we qualitatively compare event collections created from a web archive to event collections created from the live web. Second, we present a case study about European elections. Finally, we conduct a source analysis examining different sources for timeline generation and compare the performance of different LLMs in timeline generation.

6.1 The Live Web Results on Event Collections

In Section 5.3, we presented our user study of 10 events using the Portuguese Web Archive (PWA). Now, we replicate this process using the Bing Web Search API[21] to create event collections for the same events[22], allowing us to compare insights from the live web data.

[21] https://www.microsoft.com/en-us/bing/apis/bing-web-search-api

[22] https://eventexplorer.l3s.uni-hannover.de/user_study_bing_en/

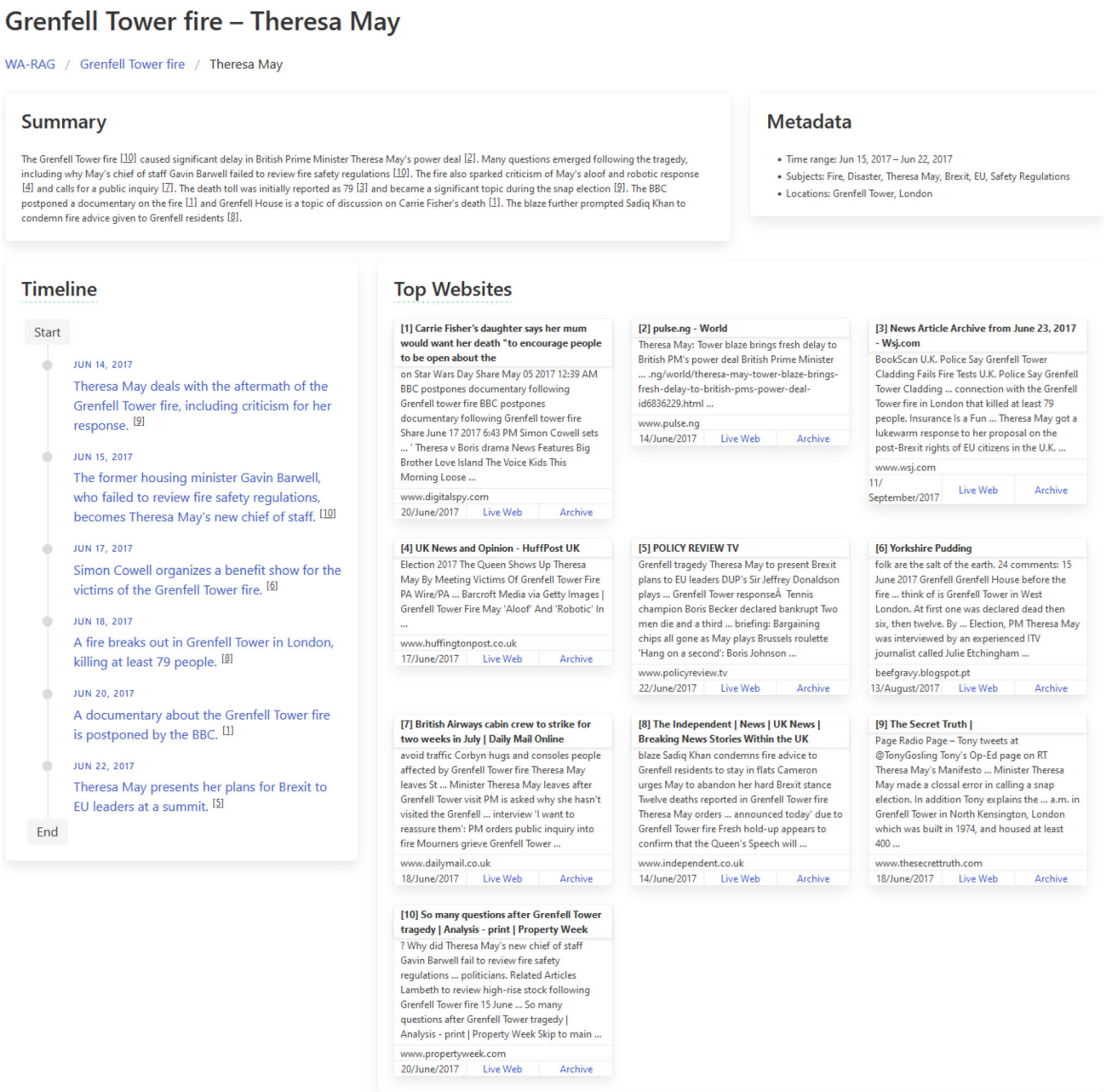


Fig. 8 Screenshot of the *EventExplorer* interface showing the event "Grenfell Tower fire" given the aspect "Who" and related entity "Theresa May" (web archive result)

Our observations indicate that the most significant difference between the collections generated using the Bing API and those using the PWA lies in the quality of the search results. The Bing API consistently returns highly relevant results, while the search functionality of web archives often provides less relevant results.

Additionally, the quality of snippets retrieved from the Bing API is better than those from the PWA. In the web archives results, snippets could come from less important parts of the websites, such as navigational elements. This misalignment impacts the generated event collections: first, the quality of created descriptive components like timelines is affected. Second, when following the presented links, users might find the website content unrelated.

In contrast, the Bing API provides highly relevant snippets that accurately represent the content of the entire website. This improves both the navigation experience – ensuring users land on relevant websites – and the generation of descriptive components, as the snippets offer a reliable summary of the website's content.

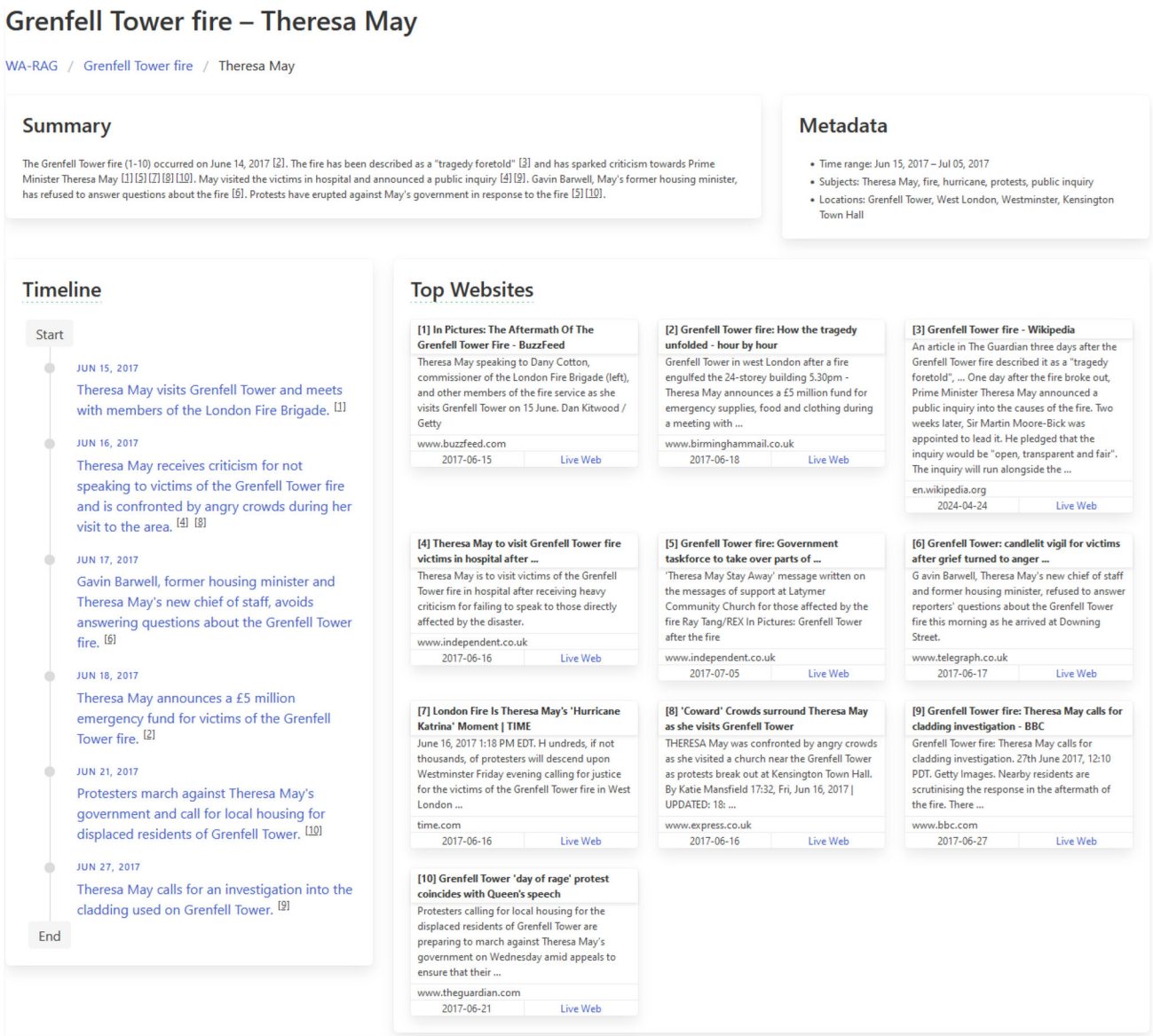


Fig. 9 Screenshot of the *EventExplorer* interface showing the event "Grenfell Tower fire" given the aspect "Who" and related entity "Theresa May" (the live web result)

As an example, Fig. 8 and Fig. 9 illustrate the event "Grenfell Tower fire" focusing on the aspect "Who" and the related entity "Theresa May" based on websites in the PWA and Bing, respectively. All timeline items in the live web version (Fig. 9) are about Theresa May. In contrast, the timeline generated from the web archive results contains six items, where three of them do not mention "Theresa May" (Fig. 8).

6.2 Case Study on Elections in Europe

We conduct a case study on event collection creation on the live web by generating event collections for a large set of European elections using the Bing API.

In this case study[23], we focus on recent elections in Europe such as the "Romanian presidential election in 2019" and the "German presidential election in 2022". With this focus, we aim to provide a selection of events that are similar in structure but diverse in their cultural perception. Additionally, we note that the analysis of elections on the web has attracted major attention in recent years. This is evidenced by Wei et al. [63] who provide an overview of 44 papers about elections on the web published between 1994 to 2023 at The Web Conference alone – revealing the impact of the web on elections through data-driven election campaigns, the spread of false information and more.

[23] https://eventexplorer.l3s.uni-hannover.de/elections_bing_en/

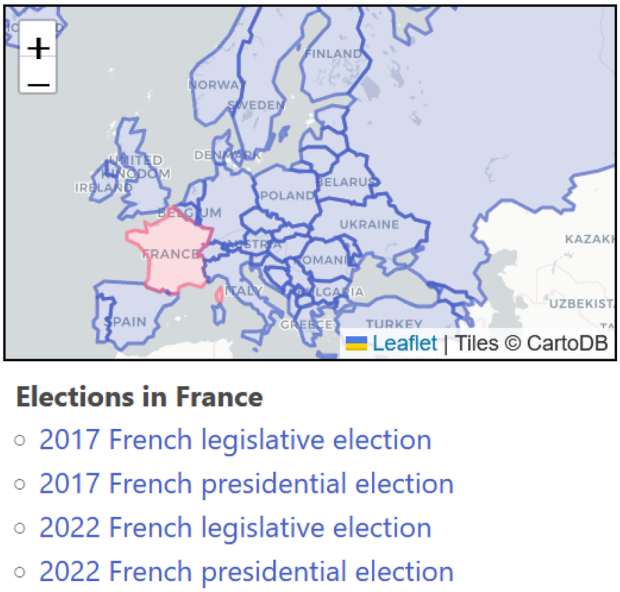


Fig. 10 Case Study of elections in Europe in *EventExplorer*

We aim to facilitate the exploration and understanding of these events across regions and through our created collections by covering a wide range of countries and electoral events. Precisely, we select the last two presidential and legislative elections from 50 European countries as our set of events in this case study[24]. This way, we retrieve a total of 166 elections in Europe for which we retrieve and rank documents and generate the descriptive components.

To follow up on *(RQ3)*, this case study is based on websites retrieved from the live web employing the Bing API. This allows us to assess the potential advantages of accessing high quality results of current search engines on creating event collections, and examine the adaptability of *EventExplorer* across different web environments.

Fig. 10 shows the main interface into *EventExplorer*'s collection of European elections which is also available online[25]. Following an analysis of the 166 elections accessible through this interface, we derive three main observations: (i) By applying *EventExplorer*'s event collection generation strategies on a large set of events using a different source, we demonstrate its generalisability. (ii) Through the event aspects retrieval with a knowledge graph, we ensure unbiased selection of connected entities, which are mainly politicians and parties who have been candidates of the specific elections. (iii) Through the web search API, we retrieve high-quality results, however, at a loss of rich archived information reflecting the web's perception of an election at the time happening.

6.3 Source Analysis

We analyse the benefits and downsides of five different sources for timeline generation *(RQ3)*. Apart from the **PWA** and **Bing** search results introduced before, we leverage news collections by utilising **Google**'s news search within the same time frame as *EventExplorer*, extracting snippets from the top 20 news articles and constructing a timeline as described in Section 4.7. Additionally, we ask Microsoft **Copilot**, a RAG-based chatbot for timeline generation [64]. Lastly, as an example of an LLM without performing RAG, we ask **ChatGPT** for timeline generation.

For our analysis, we generate five timelines of the event "Poisoning of Sergei and Yulia Skripal" using the described sources. Fig. 11 demonstrates the screenshots of the results obtained from these sources. Table 5 provides statistics of the timelines, illustrating significant differences among the sources. Since it does not use RAG, ChatGPT lacks links, and Copilot only provides 3 links. PWA using *EventExplorer* emerges as the most diverse approach, offering aspects and a robust text diversity score of 0.62 and a mean temporal deviation score of 0.49. Regarding the accuracy of the timeline dates, Bing and Google News using *EventExplorer* outperform other approaches. Notably, the cleaner snippets from Google and Bing compared to the snippets from PWA likely contribute to their higher date accuracy. Finally, while Copilot and ChatGPT cover wide time ranges of multiple years, PWA, Bing and Google clearly focus on documents close to the event's time and thus allow exploring the event from a perspective of when the event actually unfolded.

6.4 LLM Analysis

In addition to analysing different sources for timeline generation, we evaluate the performance of various LLMs in generating timelines from the same source. For this, we generate timelines given the Bing search results for the same event as in the previous section ("Poisoning of Sergei and Yulia Skripal") using five different LLMs: OpenAI's GPT-3.5 and GPT-4o[26], Anthropic's Claude 3.5 Haiku[27], as well as the smaller open-source LLMs Llama 3.3 [65] and Microsoft's Phi-14[28]. The results are summarized in Table 6 using the same metrics as in the source analysis.

[24] We selected the last legislative and presidential elections (https://en.wikipedia.org/wiki/List_of_next_general_elections) and queried EventKG for their respective previous editions.

[25] https://eventexplorer.l3s.uni-hannover.de/elections_bing_en/

[26] https://platform.openai.com/docs/models#gpt-4o

[27] https://www.anthropic.com/claude/haiku

[28] https://techcommunity.microsoft.com/blog/aiplatformblog/introducing-phi-4-microsoft's-newest-small-language-model-specializing-in-comple/4357090

Table 5 Comparative analysis of timeline generation by different sources and methods. EE-G indicates the integration of timeline generation component of *EventExplorer* on the source data

	Mean Temp.Deviation	#Links	Text Diversity	Correct Dates	Time Span
PWA + EE-G	**0.49**	16	**0.62**	60%	4 months
Bing + EE-G	0.39	**17**	0.52	**90%**	5 months
Google News + EE-G	0.35	12	0.54	89%	3 months
Copilot	**0.49**	3	0.43	33%	2 years
ChatGPT	0.38	0	0.61	20%	3 years

Table 6 Comparative analysis of timeline generation by different LLMs using Bing search results (i.e., variations of Bing + EE-G)

LLM	#LLM Parameters	Mean Temp. Deviation	#Links	Text Diversity	Correct Dates	Time Span
GPT-3.5	175B	0.39	17	0.52	90%	5 months
GPT-4o	1.8T	0.39	**20**	0.62	**100%**	4 months
Claude 3.5 Haiku	unknown	**0.46**	11	**0.72**	**100%**	4 months
Llama 3.3	70B	0.43	12	0.63	85%	4 months
Phi-4	14B	No valid JSON output generated				

GPT-4o demonstrates the best overall performance, achieving high accuracy, diversity, and link inclusion, although it falls behind the other LLMs in terms of temporal deviation. Claude 3.5 Haiku exhibits the highest text diversity but provides the fewest links to the source documents. GPT-3.5 extends the timeline's timespan but is slightly outperformed by the other LLMs in terms of text diversity and date accuracy.

Among the open-source LLMs, while Phi-4 (14B parameters) is able to understand the task of timeline generation, it was unable to generate the desired JSON output. However, it is worth noting that, despite having much less parameters (70B compared to GPT3.5's 175B), the open-source LLM LLama 3.3 still generates timelines of comparable quality. Thus, Llama 3.3 is an efficient and accessible alternative for *EventExplorer*'s timeline generation.

7 Discussion

In this section, we discuss our findings from generating event collections from web archives, highlighting our successes, challenges, and the key insights gained. Following this, we analyse the lessons learned from generating event collections from the live web and how its use addresses the challenges faced with web archives.

7.1 Insights and Challenges in Web Archives Event Collection

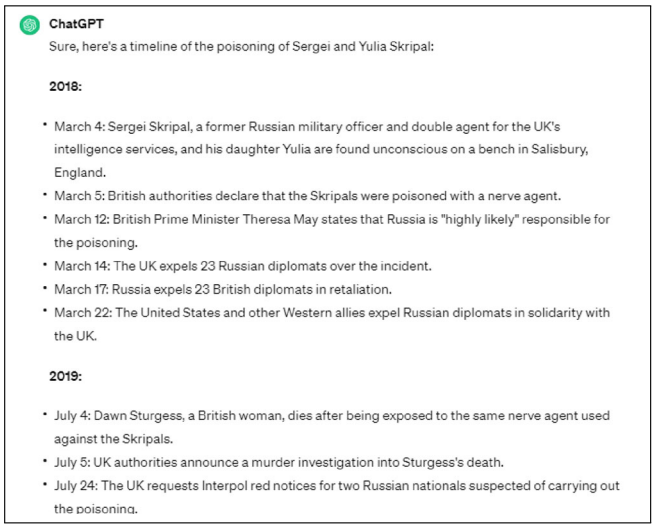
A core part of our article is the idea of creating event collections from web archives, aimed at automating tasks that web archivists usually manage through manual curation. While we encountered several challenges, the feedback from web archivists on our results has been promising.

Automated methods offer scalability and efficiency of event collection creation. These methods can potentially process large amounts of data, facilitating the creation of comprehensive event collections, and overcoming budget constraints and limitations in expert availability. Moreover, automated approaches can provide descriptive components for these collections, helping users navigate and increasing engagement.

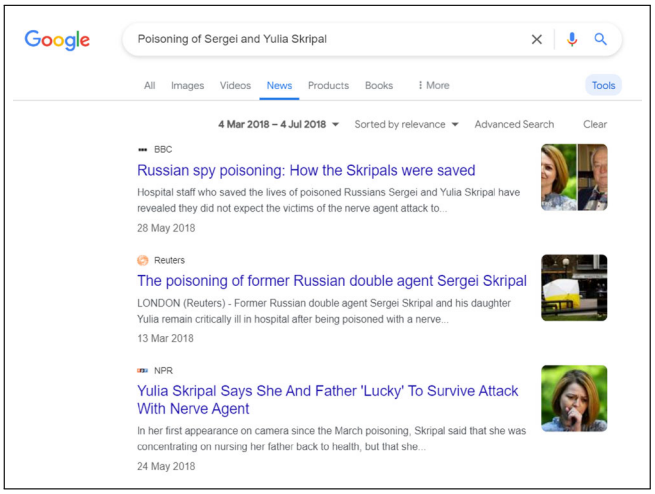
By employing approaches such as *EventExplorer*, web archivists can broaden the scope of their collections, capturing a more diverse range of events and their aspects. Notably, they value the descriptive components of our approach, which increase narrative clarity and understanding of event collection.

However, challenges arise from the different priorities between the two discussed approaches, i.e., automated curation and manual curation. While automated methods prioritise achieving efficiency and scalability in information retrieval (acknowledging the possibility of getting unrelated results as a trade-off), web archivists emphasise carefully selecting relevant websites to maintain collection integrity. This contrast highlights the need for balance and collaboration between the two groups to ensure that event collections are comprehensive and relevant and their creation is scalable and efficient *(RQ4)*.

In the following, we outline several key challenges to be addressed in future research to improve event collection creation from web archives further.

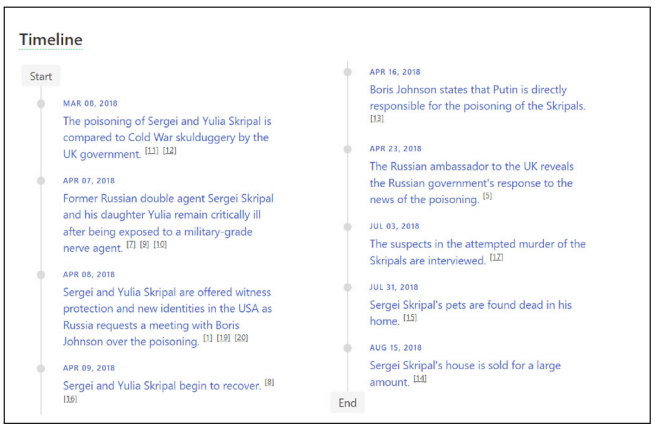


(a) ChatGPT's timeline



(b) Google News' results

(c) *EventExplorer*'s timeline using Bing



(d) *EventExplorer*'s timeline using the PWA

(e) Microsoft Copilot's timeline

Fig. 11 Screenshots of timelines generated from various sources for the event "Poisoning of Sergei and Yulia Skripal". Due to space limitations, the ChatGPT, Microsoft Copilot and Google News screenshots do not depict the complete contents

- While the precision of document selection is of immense importance, the search functionality of web archive interfaces typically falls behind the current dense document retrieval methodologies (e.g., due to the large volumes of web archives).
- The displayed snippets in web archives search results lack the quality found in search engines like Google and Bing, impacting ranking approaches primarily based on snippets.
- Creating descriptive components using RAG, occasionally presents inaccuracies in referencing input articles. This is often due to the response quality of naive RAG, which may be affected by hallucination, leading to unrelated answers to the context [66].
- Capturing the voices on the margins and people with diverse backgrounds and views as mentioned in Section 3.3 is still a major challenge that needs more effort in applying unbiased learning approaches and requires crawling techniques that deal with blind spots in web archives [67].
- Existing collection methods and most web archive crawling approaches prioritise text while overlooking other modalities such as video and social media sources. However, according to the interviews in Tables 7, 8 and 9, there is a need for diverse modalities, building connections with people, and capturing social impact.
- According to the user study results in Section 5.3.3, diverse website publications such as news media and government statements should be included in collections. Relying only on a web archive's API and its existing crawls may miss such sources.

7.2 Event Collections: Web Archives versus Live Web

To enhance event collection creation, a few targeted improvements to web archives can be highly effective *(RQ5)*. Our case study with European Elections collections using Bing API in Section 6.2 demonstrated the advantages of using the live web data for creating event collection. We observed that Bing's search quality and the relevance of its results were better than those from web archives. Specifically, the quality of snippets was noticeably higher. These improvements highlight the potential of quality of data in addressing two key challenges identified in web archives in Section 7.1: the search functionality of web archive interfaces, and limited snippet quality. By applying approaches and practices from the live web search engines such as Bing, these issues can be addressed effectively. Web archives hold significant value for creating event collections, and improving their search capabilities and snippet quality would enhance their utility. Incorporating lessons learned from the live web can help bridge the gaps and make web archives more effective for event collection purposes.

8 Conclusion & Future Work

In this article, we addressed the task of creating collections of societally impactful events from web. Through expert interviews, we gained insights into the requirements for building and utilising event collections from the perspective of web archivists *(RQ1)*. Subsequently, we developed *EventExplorer*, our method for efficiently creating event collections and helping users in exploring them. Our analysis of the user study focused on web archives results, received mainly positive feedback from the experts, specifically regarding the descriptive components created via retrieval-augmented augmentation *(RQ2)*. We extended our collections with a focused case study on elections in Europe, leveraging search functionalities on the live web. This case study demonstrates the potential of our approach to address the identified challenges and further enhance the capabilities of *EventExplorer (RQ3)*.

Our findings emphasise the importance of developing automatic collection creation approaches and highlight the significance of collaboration between information retrieval researchers and web archivists to meet the needs of users in navigating and understanding web archives content *(RQ4)*. For bringing together manual curation of web archive collections and automatic creation, we identified a set of key challenges that serve as a basis for future endeavours in the fields of information retrieval and web archives *(RQ5)*.

For future work, we see several potential directions to extend *EventExplorer*, including enabling the real-time creation of event collections. However, we should consider constraints such as the time required for collecting and processing data and the costs associated with APIs, such as OpenAI, which will be critical factors in the scalability and practicality of our solution. Another potential direction is incorporating functionalities that allow web archivists to modify and refine collected event collections, providing flexibility in curating and maintaining collections. Additionally, interviews with experts from diverse backgrounds, including those from various web archiving institutions will help to gain a more comprehensive understanding of user and archivist needs.

A Appendix

Tables 7, 8 and 9 summarize our interviews conducted as described in Section 3.

Table 7 Selected and summarised questions and answers of our expert interviews (part 1)

Question	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
How do you decide for which topics you create collections?	Significant aspects of a country such as elections, sports events and unfolding events such as natural disasters	Topics with public interest	Based on a scoping document	Important events to a country like elections and sporting events	Social history of a country such as elections
How are the topics for sub-collections selected?	Determined by the curator who is approaching the subject	-	Significant sub-events of the main event	Breaking a big collection down to different aspects such as business and society if possible	Breaking down big collections when the number of websites increases
Do you have any criteria for selecting website domains?	Archiving only national websites and considering privacy issues for social media	Popularity of a website and its relevance to the topic	Academia, government, social media and civil society	Trustworthiness of the websites, news media and social media	Collecting public content as broadly as possible
What is your process of creating a collection?	According to the scoping document which describes the goal of a collection, the curator's background, size and time span	Defining the scope, starting point (seed URL) for the target community and finally dissemination of the results	Determined in the scoping document	Defining a plan including criteria, size, time, number of staff, topics and sub-topics	Defining a scoping document which states the subject, areas to cover, sub-categories, public communications and dates
What expertise is required to create a collection?	Subject expertise, having contacts with community groups and technical experience with web archiving (good but not necessary)	In-depth knowledge of the topic, having contact with the related community, generic knowledge of web content and knowing at least one content recording tool	Content, web archiving and institutional knowledge i.e, understanding how these projects work	Fair knowledge of what happens in a country (for national library institutions)	A person with a background in library science who can handle metadata
Why do you create collections?	To be more inclusive and representative: archiving the websites and the voices of unrepresented people.	-	-	Due to the legal requirement for preserving digital publications in a national repository	Witnessing current time through websites which are an important record of our time

Table 8 Selected and summarised questions and answers of our expert interviews (part 2)

Question	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
Who are the users of these collections?	General public, academic researchers, journalists, humanists, sociologists and data scientists	Researchers and common citizens	We don't know exactly, maybe researchers	Not known	General public
For which purpose are users using the collections?	Conducting research, looking for a document or a particular statement and looking at web archives as data	Looking for information and documents in the past	-	Looking for a specific document	Looking for old documents or expecting the archived website to behave as on live web
How can an event collection be different from non-event collections?	Difference in choosing the time span of a collection due to the temporal aspect of events	Temporal aspect of events and limited time to collect the websites	Easier scoping of event collections due to temporal and spatial aspects and clear boundaries	Temporal aspect of event collections and limited time to collect websites	Temporal aspect of event collections, limited time in interest and explosion of publications over this time span
Are there different user information needs for event vs. non-event collections?	Event collections are more self-explanatory while non-event collections need more contextual information	Small collections also have small communities or specific communities	-	-	-
Do you provide a single snapshot per website in a collection?	Depends on the topic. Mostly several snapshots	Multiple snapshots	Multiple snapshots	Depending on the topic and the type of the website. Some are captured daily, some weekly	
How is the size of a collection determined?	Largest collection: 8000. The minimum would be 50	10 megabytes for each website. But no limit on the number of websites	No limit	No limit	Minimum 50 websites. 1000 is pretty reasonable if it's broken down into sub-categories

Table 9 Selected and summarised questions and answers of our expert interviews (part 3)

Question	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
What metadata do you collect in a collection?	URL and timestamps, descriptions, domain extension, the crawling year and taxonomy of subject terms	Description and title of the website	URL, title, description and type of content (media, text, ...)	Description, URL, subject tag, name of the organisation, date and evidence that the website is national	Title, description and the dates
How is the time span for an event collection selected?	As long as the event is relevant and material is being produced about it	Different depending on the collection, the community and the content	For some events half a year, even one year before preparation started	One to two months after the events such as the General elections and world cup	As long as the event is a relevant topic
Do event collections focus on specific aspects or aim for comprehensive representation?	Our ambition is to be comprehensive	A comprehensive representation	It is impossible because we have to capture what is significant	We try to put all aspects into it	It depends on the scoping of the collection and the determined areas that would be important to collect
Are the collection characteristics different across different kinds of events?	Criteria for creating the collection are the same across all event types	-	No difference	-	
Have you ever used already captured websites or do you crawl yourself?	It is a mixture of both currently using crawled websites and crawling	-	Not yet	We have used already archived websites	We have done it before
What is a good collection?	-	-	A collection with diverse websites, metadata and video capable of answering research questions	A collection focusing on all aspects, including environment, economic, social outcomes and impacts	A collection with a real connection to people and what is going on in a country, with acceptable size and presentation

Acknowledgements We would like to thank the following web archive experts and curators for their support in the interviews and user study: Ricardo Basilio (The Portuguese Web Archive), Aled Betts (The National Library of Wales), and Nicola Bingham (The British Library).

Author Contributions All authors contributed to the conceptualisation. Interviews, methodology, implementation and experiments were performed by Sara Abdollahi. Website development was performed by Simon Gottschalk. Supervision of Sara Abdollahi was done by Wolfgang Nejdl and Simon Gottschalk. The first draft of the manuscript was written by Sara Abdollahi and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Materials availability The *EventExplorer* website is available on https://eventexplorer.l3s.uni-hannover.de/.

Code Availability Our code is available on GitHub: https://github.com/saraabdollahi/EventExplorer.

Declarations

Financial or non-financial interests The authors have no relevant financial or non-financial interests to disclose.

Conflict of interest/Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval and consent to participate All participants of our interview and user study voluntarily agreed to participate and provided written informed consent.(Upon publication, we will ask the interviewees if they want to be acknowledged by name at the end of the article.)

References

1. Demidova, E., Risse, T.: "Creating event-centric collections from web archives," in *The Past Web: Exploring Web Archives*. Springer, pp. 57–67. (2021)
2. Risse, T., Demidova, E., Gossen, G.: "What do you want to collect from the web," in *Proceedings of the Building Web Observatories Workshop (BWOW)*, vol. 2014. Citeseer, (2014)
3. Jayanetti, H.R., Jones, S.M., Klein, M., Osbourne, A., Koerbin, P., Nelson, M.L., Weigle, M.C.: "Creating structure in web archives with collections: Different concepts from web archivists," in *International Conference on Theory and Practice of Digital Libraries*. Springer, pp. 450–458. (2022)
4. Gossen, G., Demidova, E., Risse, T.: "Analyzing web archives through topic and event-focused sub-collections," in *Proceedings of the 2016 ACM Conference on Web Science*, pp. 291–295. (2016)
5. AlNoamany, Y., Weigle, M.C., Nelson, M.L.: "Generating stories from archived collections," in *Proceedings of the 2017 ACM Conference on Web Science*, pp. 309–318. (2017)
6. Holzmann, H., Nejdl, W.: "A holistic view on web archives," in *The Past Web: Exploring Web Archives*. Springer, pp. 85–99. (2021)
7. Wang, X., Xie, Z.: "The case for alternative web archival formats to expedite the data-to-insight cycle," in *2020 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pp. 177–186. (2020)
8. Costa, M.: "Full-text and URL search over web archives," in *The past web: Exploring Web Archives*. Springer, pp. 71–84. (2021)
9. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., et al.: Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems **33**, 9459–9474 (2020)
10. Costa, M., Silva, M.J.: "Understanding the information needs of web archive users," in *Proceedings of the 10th International Web Archiving Workshop*, vol. 9, , p. 6. (2010)
11. Ogden, J., Halford, S., Carr, L.: "Observing web archives: The case for an ethnographic study of web archiving," in *Proceedings of the 2017 conference ACM on Web Science*, pp. 299–308. (2017)
12. Hockx-Yu, H.: Access and scholarly use of web archives. Alexandria **25**(1–2), 113–127 (2014)
13. Jones, S.M., Nwala, A., Weigle, M.C., Nelson, M.L.: "The many shapes of archive-it," *arXiv preprint* arXiv:1806.06878, (2018)
14. Jones, S., Klein, M., Weigle, M., et al.: "MementoEmbed and Raintale for web archive storytelling," *arXiv preprint* arXiv:2008.00137, 2021
15. Jones, S.M., Weigle, M.C., Klein, M., Nelson, M.L.: "Hypercane: Intelligent sampling for web archive collections," in *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, pp. 316–317. (2021)
16. Jones, S.M., Weigle, M.C., Nelson, M.L.: Hypercane: Toolkit for summarizing large collections of archived webpages. ACM SIGWEB Newsletter **2021**(Summer), 1–14 (2021)
17. Jones, S., Weigle, M., Nelson, M.: "Social cards probably provide for better understanding of web archive collections," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 2023–2032. (2019)
18. Jones, S.M.: "Improving collection understanding in web archives," *Bull. IEEE Tech. Comm. Digit. Libr.*, vol. 15, no. 1, (2019)
19. Ruest, N., Fritz, S., Deschamps, R., Lin, J., Milligan, I.: From archive to analysis: Accessing web archives at scale through a cloud-based interface. International Journal of Digital Humanities **2**(1), 5–24 (2021)
20. Kelly, M., Brunelle, J.F., Weigle, M.C., Nelson, M.L.: "A method for identifying personalized representations in web archives," *D-Lib Magazine*, vol. 19, no. 11-12, (2013)
21. Campos, R., Correia, D., Jatowt, A.: "Public News Archive: A searchable sub-archive to Portuguese past news articles," in *European Conference on Information Retrieval*. Springer, pp. 211–216. (2023)
22. Nanni, F., Ponzetto, S.P., Dietz, L.: "Building entity-centric event collections," in *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, pp. 1–10. (2017)
23. Gossen, G., Risse, T., Demidova, E.: Towards extracting event-centric collections from web archives. International Journal on Digital Libraries **21**(1), 31–45 (2020)

24. Plötzky, F., Kiehne, N., Balke, W.-T.: "Lost in recursion: Mining rich event semantics in knowledge graphs," in *Proceedings of the 2024 ACM Web Science Conference*, pp. 354–364. (2024)
25. Nogueira, R., Cho, K.: "Passage re-ranking with BERT," *arXiv preprint* arXiv:1901.04085, (2019)
26. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint* arXiv:1810.04805, (2018)
27. Hofstätter, S., Althammer, S., Schröder, M., Sertkan, M., Hanbury, A.: "Improving efficient neural ranking models with cross-architecture knowledge distillation," *arXiv preprint* arXiv:2010.02666, (2020)
28. Hofstätter, S., Lin, S.-C., Yang, J.-H., Lin, J., Hanbury, A.: "Efficiently teaching an effective dense retriever with balanced topic aware sampling," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 113–122. (2021)
29. Li, C., Yates, A., MacAvaney, S., He, B., Sun, Y.: PARADE: Passage representation aggregation for document re-ranking. ACM Transactions on Information Systems **42**(2), 1–26 (2023)
30. Nogueira, R., Yang, W., Cho, K., Lin, J.: "Multi-stage document ranking with BERT," *arXiv preprint* arXiv:1910.14424, (2019)
31. Dalton, J., Dietz, L., Allan, J.: "Entity query feature expansion using knowledge base links," in *Proceedings of the 37th international ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 365–374. (2014)
32. Xiong, C., Callan, J.: "Query expansion with Freebase," in *Proceedings of the 2015 International Conference on the Theory of Information Retrieval*, pp. 111–120. (2015)
33. Dahir, S., El Qadi, A.: A query expansion method based on topic modeling and DBpedia features. International Journal of Information Management Data Insights **1**(2), 100043 (2021)
34. Zong, N., Lee, S., Kim, H.-G.: Discovering expansion entities for keyword-based entity search in linked data. Journal of Information Science **41**(2), 209–227 (2015)
35. Jain, S., Seeja, K., Jindal, R.: A fuzzy ontology framework in information retrieval using semantic query expansion. International Journal of Information Management Data Insights **1**(1), 100009 (2021)
36. Rosin, G.D., Guy, I., Radinsky, K.: "Event-driven query expansion," in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pp. 391–399. (2021)
37. Abdollahi, S., Kuculo, T., Gottschalk, S.: "Event-specific document ranking through multi-stage query expansion using knowledge graphs," in *European Conference on Information Retrieval*. Springer, (2024)
38. Gottschalk, S., Demidova, E.: EventKG-the Hub of event knowledge on the web-and biographical timeline generation. Semantic Web **10**(6), 1039–1070 (2019)
39. Guu, K., Lee, K., Tung, Z., Pasupat, P., Chang, M.: "Retrieval augmented language model pre-training," in *International Conference on Machine Learning*. PMLR, pp. 3929–3938. (2020)
40. Lewis, M., Ghazvininejad, M., Ghosh, G., Aghajanyan, A., Wang, S., Zettlemoyer, L.: "Pre-training via paraphrasing," *Advances in Neural Information Processing Systems*, vol. 33, pp. 18 470–18 481, (2020)
41. Hossain, N., Ghazvininejad, M., Zettlemoyer, L.: "Simple and effective retrieve-edit-rerank text generation," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 2532–2538. (2020)
42. Kazemnejad, A., Salehi, M., Baghshah, M.S.: "Paraphrase generation by learning how to edit from samples," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 6010–6021. (2020)
43. Pride, D., Cancellieri, M., Knoth, P.: "CORE-GPT: Combining open access research and large language models for credible, trustworthy question answering," in *International Conference on Theory and Practice of Digital Libraries*. Springer, pp. 146–159. (2023)
44. Ji, Y., Li, Z., Meng, R., Sivarajkumar, S., Wang, Y., Yu, Z., Ji, H., Han, Y., Zeng, H., He, D.: "RAG-RLRC-LaySum at BioLaySumm: Integrating Retrieval-Augmented Generation and Readability Control for Layman Summarization of Biomedical Texts," *arXiv preprint* arXiv:2405.13179, (2024)
45. He, J., Tallam, S., Nakshathri, S., Amarnath, N., Kr, P., Kumar, D.: "Infrrd. ai at SemEval-2024 Task 7: RAG-based end-to-end training to generate headlines and numbers," in *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pp. 940–951. (2024)
46. Zhang, L., Zhang, H., Wang, C., Liang, P.: "RAG-Enhanced Commit Message Generation," *arXiv preprint* arXiv:2406.05514, (2024)
47. Quraishi, M., Fafalios, P., Herder, E.: "Viewpoint discovery and understanding in social networks," in *Proceedings of the 2018 ACM conference on web science*, pp. 47–56. (2018)
48. Draws, T., Roy, N., Inel, O., Rieger, A., Hada, R., Yalcin, M.O., Timmermans, B., Tintarev, N.: "Viewpoint diversity in search results," in *European Conference on Information Retrieval*. Springer, pp. 279–297. (2023)
49. Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G.D., Gutierrez, C., Kirrane, S., Gayo, J.E.L., Navigli, R., Neumaier, S., et al.: Knowledge graphs. ACM Computing Surveys (Csur) **54**(4), 1–37 (2021)
50. Gomes, D., Cruz, D., Miranda, J., Costa, M., Fontes, S.: "Search the past with the Portuguese web archive," in *Proceedings of the 22nd International Conference on World Wide Web*, pp. 321–324. (2013)
51. Clarke, C.L., Kolla, M., Cormack, G.V., Vechtomova, O., Ashkan, A., Büttcher, S., MacKinnon, I.: "Novelty and diversity in information retrieval evaluation," in *Proceedings of the 31st annual international ACM SIGIR conference on Research and Development in Information Retrieval*, , pp. 659–666. (2008)
52. Piryani, R., Aussenac-Gilles, N., Hernandez, N.: "Comprehensive survey on ontologies about event," in *ESWC Workshops on Semantic Methods for Events and Stories (SEMMES@ ESWC 2023)*, vol. 3443. ceur-ws. org, pp. 1–15. (2023)
53. Van Hage, W.R., Malaisé, V., Segers, R., Hollink, L., Schreiber, G.: Design and use of the simple event model (SEM). Journal of Web Semantics **9**(2), 128–136 (2011)
54. Lin, J., Nogueira, R., Yates, A.: *Pretrained transformers for text ranking: BERT and beyond*. Springer Nature, (2022)
55. Khattab, O., Zaharia, M.: "ColBERT: Efficient and effective passage search via contextualized late interaction over BERT," in *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, pp. 39–48. (2020)
56. Demidova, E., Fankhauser, P., Zhou, X., Nejdl, W.: "DivQ: Diversification for keyword search over structured databases," in *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 331–338. (2010)
57. Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G.: Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys **55**(9), 1–35 (2023)
58. Kedzie, C., McKeown, K., Diaz, F.: "Summarizing disasters over time," in *Proc. Bloomberg Workshop on Social Good (with SIGKDD)*, (2014)
59. Hofstätter, S., Mitra, B., Zamani, H., Craswell, N., Hanbury, A.: "Intra-document cascading: Learning to select passages for neural document ranking," in *Proceedings of the 44th International ACM*

SIGIR Conference on Research and Development in Information Retrieval, pp. 1349–1358. (2021)
60. Sanh, V., Debut, L., Chaumond, J., Wolf, T.: "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," *arXiv preprint* arXiv:1910.01108, (2019)
61. Nguyen, T., Rosenberg, M., Song, X., Gao, J., Tiwary, S., Majumder, R., Deng, L.: "MS MARCO: A human generated machine reading comprehension dataset," in *Proceedings of the Workshop on Cognitive Computation: Integrating neural and symbolic approaches 2016*, vol. 2640, p. 660. (2016)
62. Fleiss, J.L.: Measuring nominal scale agreement among many raters. Psychological Bulletin **76**(5), 378 (1971)
63. Wei, Z., Xu, X., Hui, P.: Digital democracy at crossroads: A meta-analysis of web and ai influence on global elections. Companion Proceedings of the ACM on Web Conference **2024**, 1126–1129 (2024)
64. White, R.W.: "Tasks, copilots, and the future of search: A keynote at SIGIR 2023," in *ACM SIGIR Forum*, vol. 57, no. 2. ACM New York, NY, USA, pp. 1–8. (2024)
65. Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Yang, A., Fan, A., et al.: "The llama 3 herd of models," *arXiv preprint* arXiv:2407.21783, (2024)
66. Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, H.: "Retrieval-augmented generation for large language models: A survey," *arXiv preprint* arXiv:2312.10997, (2023)
67. Donig, S., Eckl, M., Gassner, S., Rehbein, M.: "Web archive analytics: Blind spots and silences in distant readings of the archived web," *Digital Scholarship in the Humanities*, p. fqad014, (2023)